Power Systems

IBM

Performance Characteristics of the POWER8™ Processor

Alex Mericas
Systems Performance
IBM Systems & Technology Group Development

POWER8

Designed for Big Data - optimized for analytics performance

Processors
flexible, fast execution of analytics algorithms

Memory
large, fast workspace to maximize business insight

Data Bandwidth
bring massive amounts of information to compute resources in real-time

Optimized for a broad range of data and analytics:

Cognos SPSS

IBM Predictive Customer Intelligence

Industry Solutions

POWER8 Processor

Technology

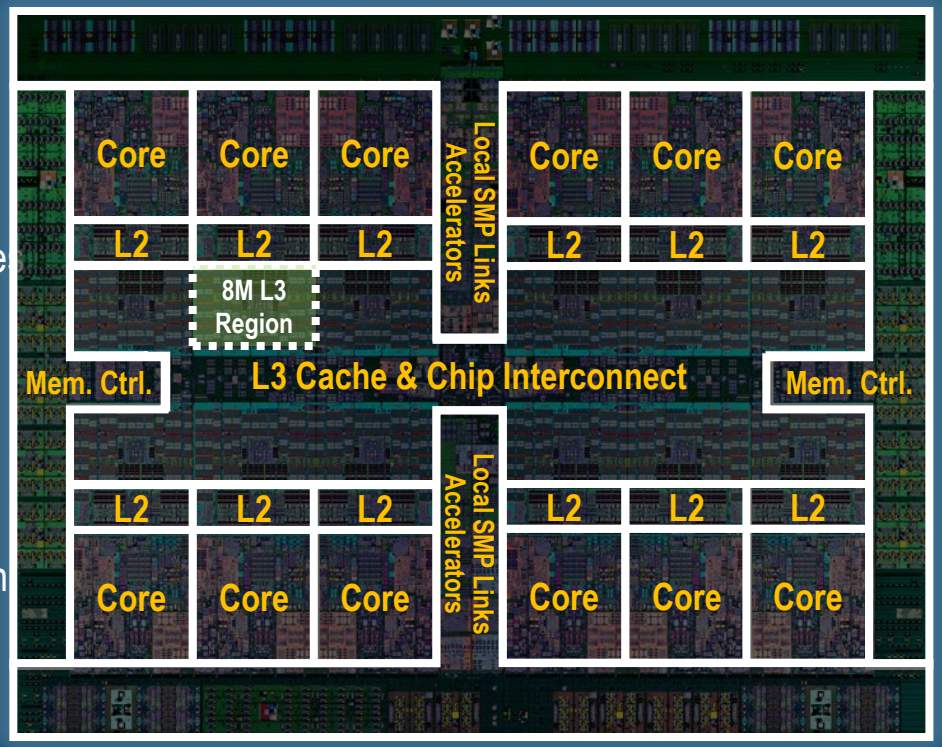
- 22nm SOI, eDRAM, 15 ML 650mm2

Cores

- 12 cores (SMT8)
- 8 dispatch, 10 issue, 16 exec pipe
- 2X internal data flows/queues
- Enhanced prefetching
- 64K data cache, 32K instruction cache

Accelerators

- Crypto & memory expansion
- Transactional Memory
- VMM assist
- Data Move / VM Mobility



Energy Management

- On-chip Power Management Micro-controller
- Integrated Per-core VRM
- Critical Path Monitors

Caches

- 512 KB SRAM L2 / core
- 96 MB eDRAM shared L3
- Up to 128 MB eDRAM L4 (off-chip)

Memory

- Up to 230 GB/s sustained bandwidth

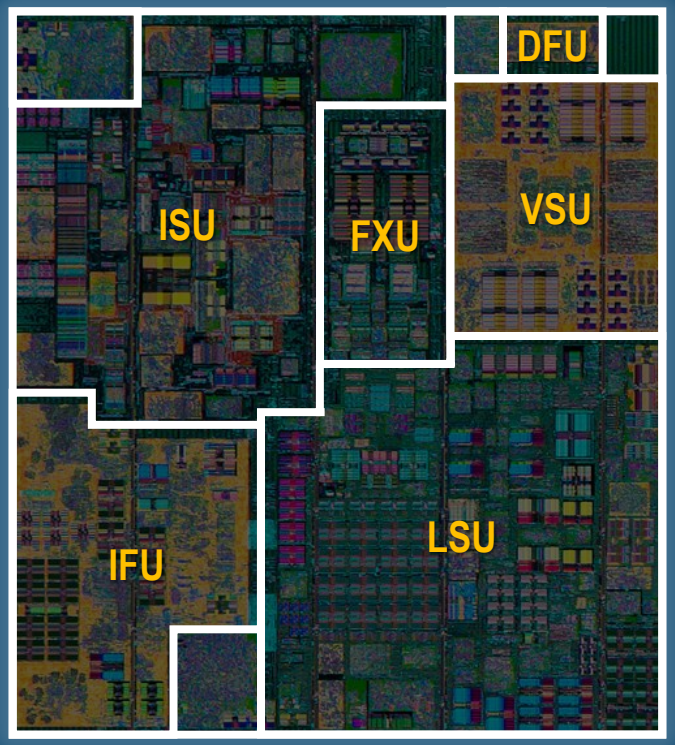
Bus Interfaces

- Durable open memory attach interface
- Integrated PCIe Gen3
- SMP Interconnect
- CAPI (Coherent Accelerator Processor Interface)

POWER8 Core

Execution Improvement vs. POWER7

- SMT4 → SMT8
- 8 instruction dispatch
- 10 instruction issue
- 16 execution pipes:
  - 2 Fixed Point, 2 Ld/Store , 2 Ld
  - 4 Floating Point, 2 Vector
  - 1 Crypto, 1 Decimal Floating Point
  - 1 Conditional, 1 Branch
- Larger Issue queues (4 x 16-entry)
- Larger completion table (28 groups)
- Larger Ld/Store reorder (128 / thrd)
- Improved branch prediction
- Improved unaligned storage access



Larger Caching Structures vs. POWER7

- 2x L1 data cache (64 KB)
- 2x outstanding data cache misses
- 4x translation Cache

Wider Load/Store

- 32B → 64B L2 to L1 data bus
- 2x data cache to execution dataflow

Enhanced Prefetch

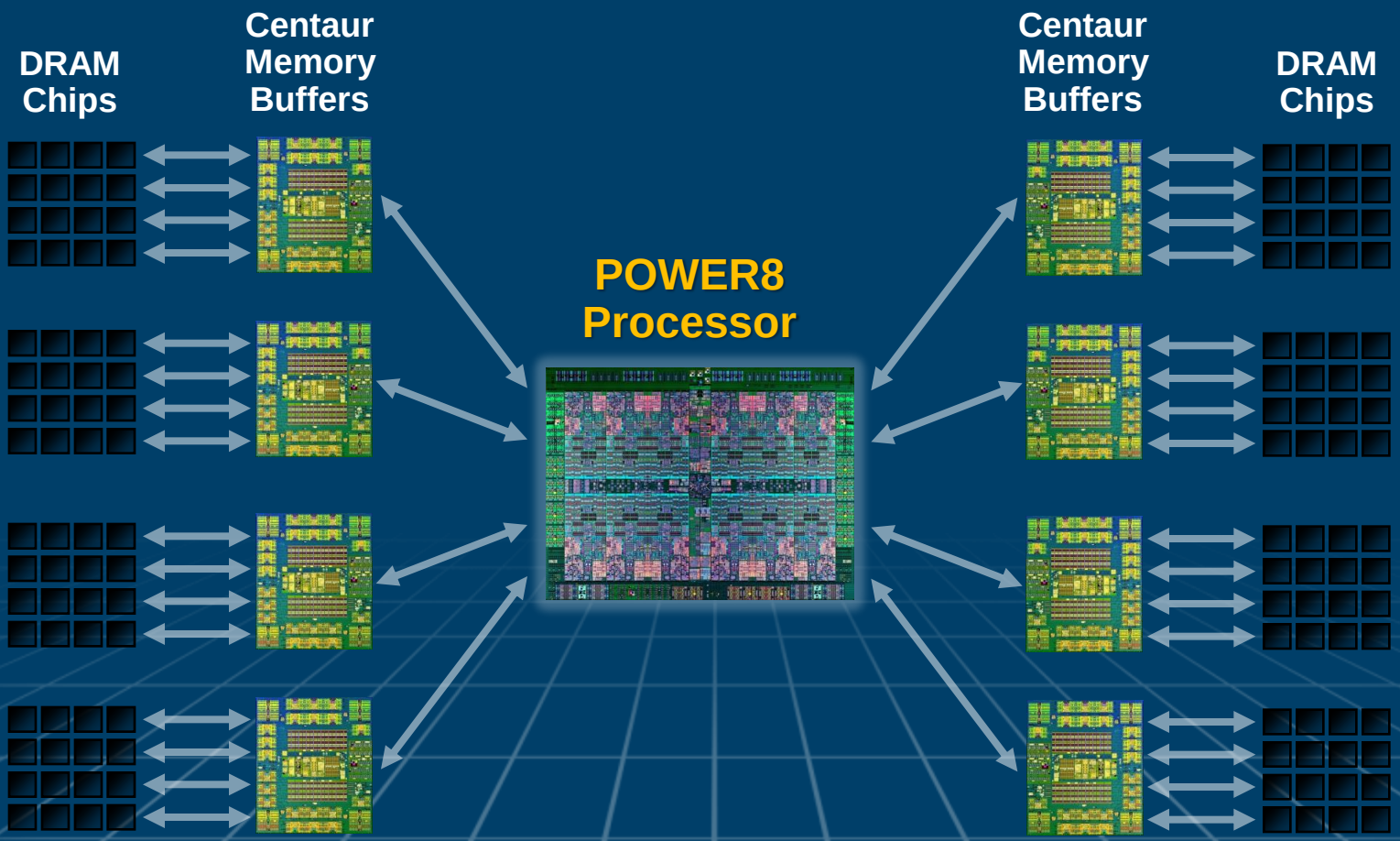
- Instruction speculation awareness
- Data prefetch depth awareness
- Adaptive bandwidth awareness
- Topology awareness

Core Performance vs . POWER7

~1.6x Thread

~2x Max SMT

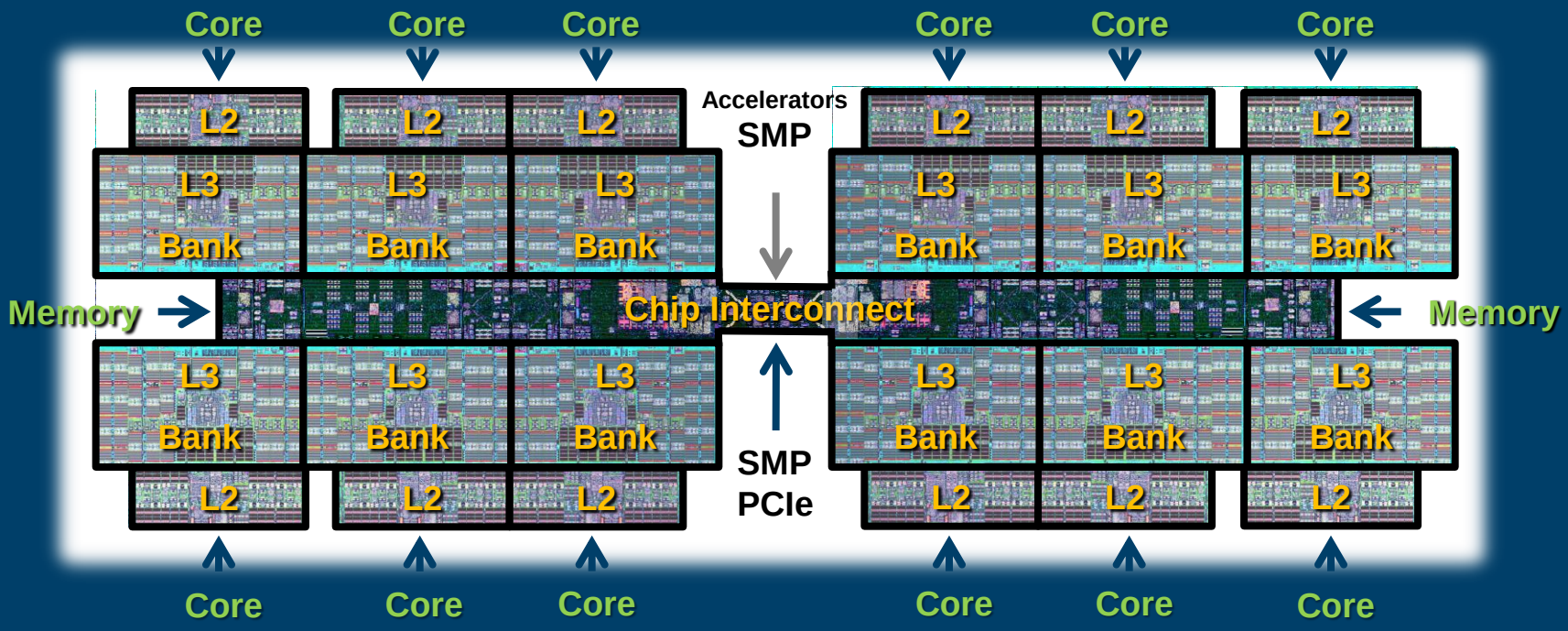
POWER8 Memory Organization



- *Up to 8 high speed channels, each running up to 9.6 Gb/s for up to 230 GB/s sustained*
- *Up to 32 total DDR ports yielding 410 GB/s peak at the DRAM*
- *Up to 1 TB memory capacity per fully configured processor socket (at initial launch)*

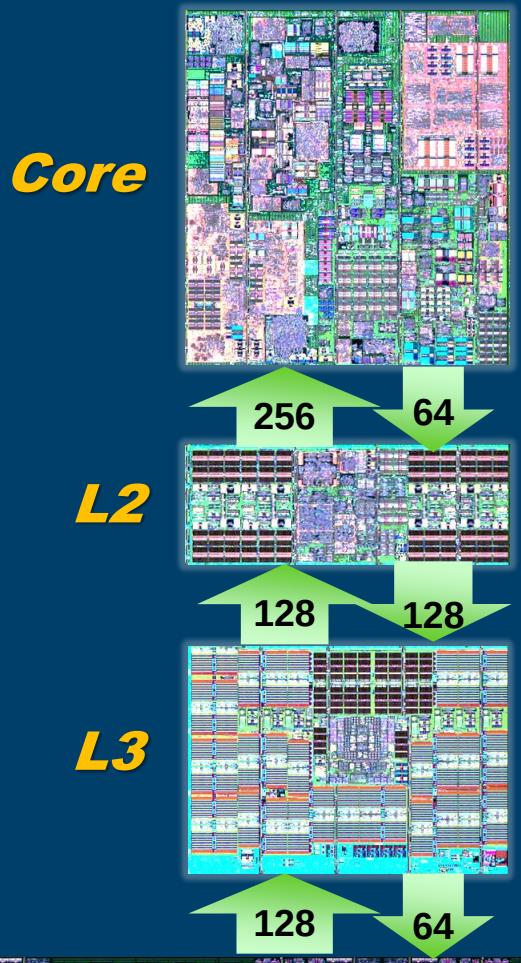
POWER8 On Chip Caches

- **L2: 512 KB 8 way per core**
- **L3: 96 MB (12 x 8 MB 8 way Bank)**
- **"NUCA" Cache policy (**Non-Uniform Cache Architecture)
  - Scalable bandwidth and latency
  - Migrate "hot" lines to local L2, then local L3 (replicate L2 contained footprint)
- **Chip Interconnect: 150 GB/sec x 16 segment per direction per segment**

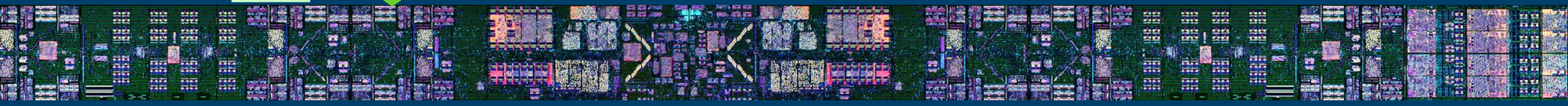


IBM

Cache Bandwidths

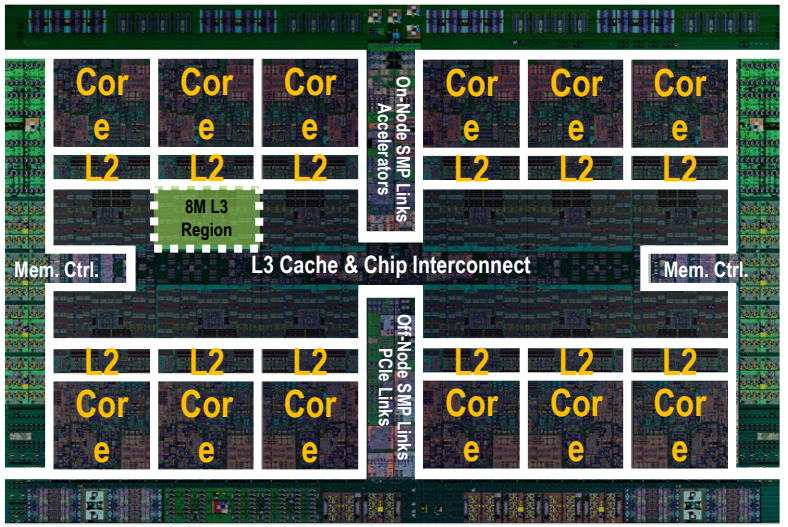


- **GB/sec shown assuming 4 GHz**
  - Product frequency will vary based on model type
- **Across 12 core chip**
  - 4 TB/sec L2 BW
  - 3 TB/sec L3 BW

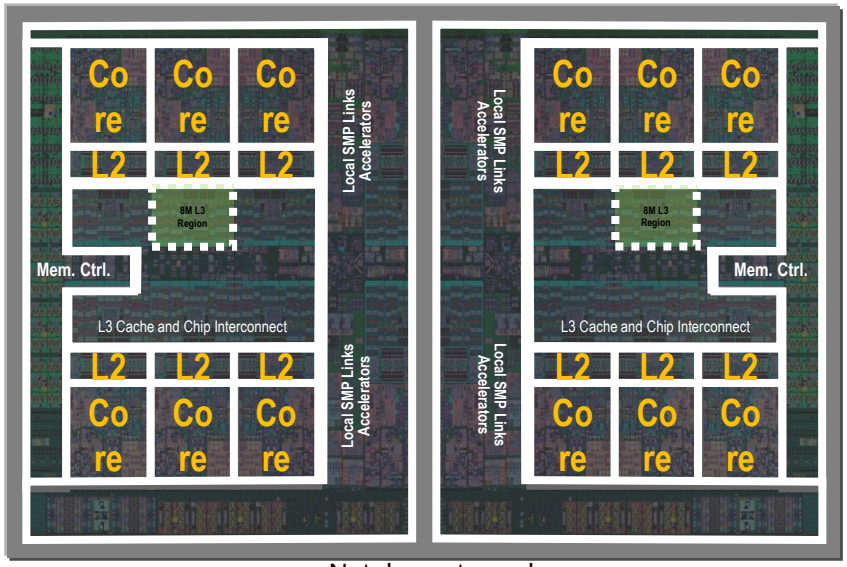


Scale-Out Processor Version (Announced April 2014)

- Scale-UP Processor (Shown at Hot Chips 25)
  - Optimized for Large SMP
  - 22nm SOI, eDRAM, 15 ML 650mm2
  - 12 Core Chip
  - 32x PCIe Gen3 (16x CAPI)
  - Large memory capacity and bandwidth



- Scale-Out Processor (1 module per socket)
  - Optimized for Scale-OUT systems
  - 2 x 6-Core Chip (362mm2 each)
  - 48x PCIe Gen3 (32x CAPI)
  - Same core, L2, L3, etc



Not drawn to scale

New Power Scale-out systems built with open innovation to put data to work

Designed for Big Data

Superior Cloud Economics

Open Innovation Platform

OpenPOWER™

Linux

Power S822L

Power S812L

- 1 or 2 sockets
- 10 or 12 cores/socket

PowerVC PowerKVM PowerVM

openstack CLOUD SOFTWARE

ubuntu Supported by Canonical redhat. SUSE

AIX i for Business Linux

Power S824 or Power S814

Power S822

- 1 or 2 sockets
- 6, 8,10 or 12 cores/socket

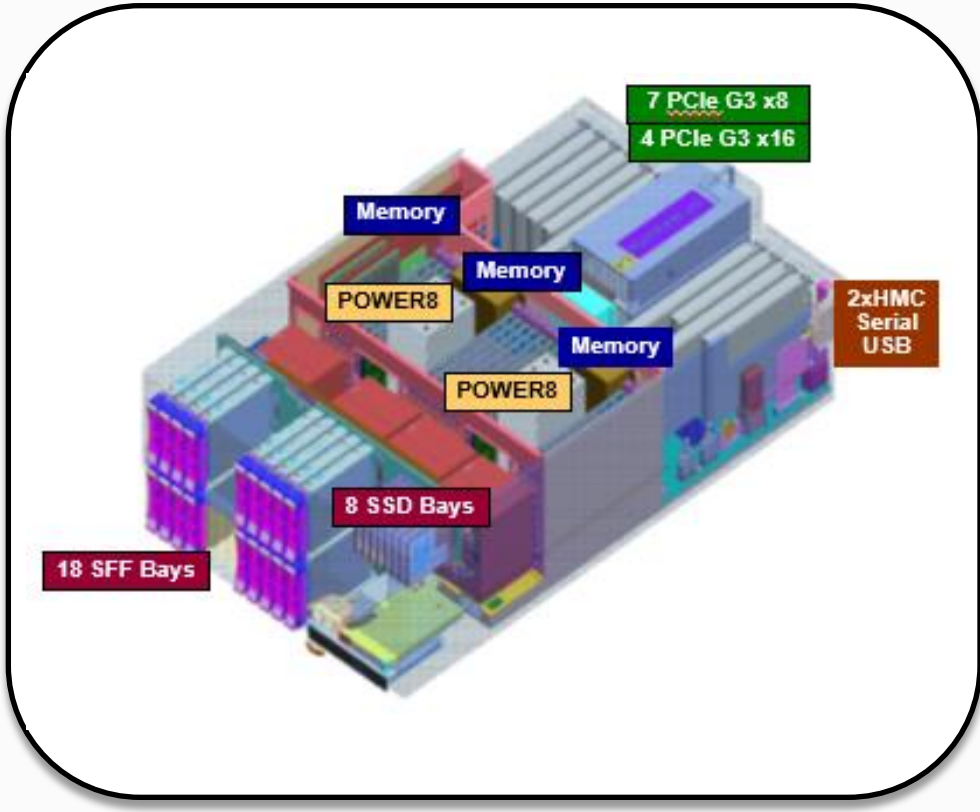
PowerVC PowerVM

openstack CLOUD SOFTWARE

redhat. SUSE

New Power Scale-out systems detailed features

- **2 Sockets** (1 socket upgradeable)
- Up to **24 cores** (192 threads)
- Up to **1 TB memory** capacity
- Hot Plug PCIe gen 3 Slots
- SR-IOV support (statement of direction)
- Ethernet: Quad 1 Gbt / (x8 slot)
- Native I/O
  - USB (3), Serial (2), HMC (2)
- Internal Storage
  - Up to 18 SFF Bays
  - Up to 8 1.8” SSD Bays (Easy Tier)
  - DVD
- Power Supplies: (200-240 AVC)



POWER8 Performance Characteristics

POWER8 CPI Stack

- Introduced with PowerPC970, the CPI stack uniquely identifies components of CPI (Cycles Per Instruction)
- Enhanced every generation to add detail and eliminate “other” category
- POWER8 splits dependency chains within a group to separate cause and effect (e.g. long latency load feeding 1 cycle add)
- Items in blue are new with POWER8

Cycles	Stalled Cycles	Stall due to BR or CR	Stall due to Branch		
			Stall due to CR		
		Stall due to Fixed Point	Stall due to Fixed-Point long		
			Stall due to Fixed-Point (other)		
		Stall due to Vector/Scalar	Stall due to Vector	Stall due to Vector long	
				Stall due to Vector (other)	
			Stall due to Scalar	Stall due to Scalar long	
				Stall due to Scalar (other)	
			Stall due to Vector/Scalar (other)		
		Stall due to Load/Store	Stall due to Dcache Miss	Stall due to L2/L3 Hit	L2/L3 hit with conflict
					L2/L3 hit with no conflict
				Stall due to L3 Miss	Stall due to On-chip L2/L3
					Stall due to On-chip Memory
					Stall due to Off-chip L2/L3
					Stall due to Off-chip Memory
					Stall due to Off-node Memory
			Stall due to LSU Reject	Reject due to Load-Hit	
				Reject due to ERAT Miss	
				Reject due to LMQ Full	
				Reject due to Reject (other)	
			Stall due to Store Finish		
			Stall due to Load Finish		
			Stall due to Store Forward		
			Stall due to Load/Store (other)		
		Stall due to Next-To-Complete Flush			
		Stall Cycles (other)			
	Waiting to Complete				
	Thread Blocked	Blocked due to LWSYNC			
		Blocked due to HWSYNC			
		Blocked due to ECC Delay			
		Blocked due to Flush			
		Blocked due to COQ Full			
		Thread Blocked (Other)			
	Completion Table Empty	Completion Table Empty due to Icache Miss	Completion Table Empty due to Icache L3 Miss		
			Completion Table Empty due to Icache Miss (Other)		
		Completion Table Empty due to Branch Mispredict			
		Completion Table Empty due to Branch Mispredict + Icache Miss			
		Completion Table Empty - Dispatch Held	Dispatch Held due to Mapper		
			Dispatch Held due to Store Queue		
			Dispatch Held due to Issue Queue		
			Dispatch Held (Other)		
		Completion Table Empty (Other)			
	Other				
	Completion Cycles				

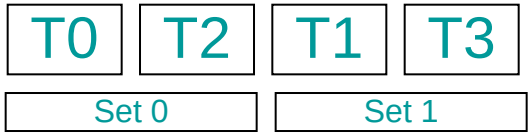
Sampled Instruction Event Register (SIER)

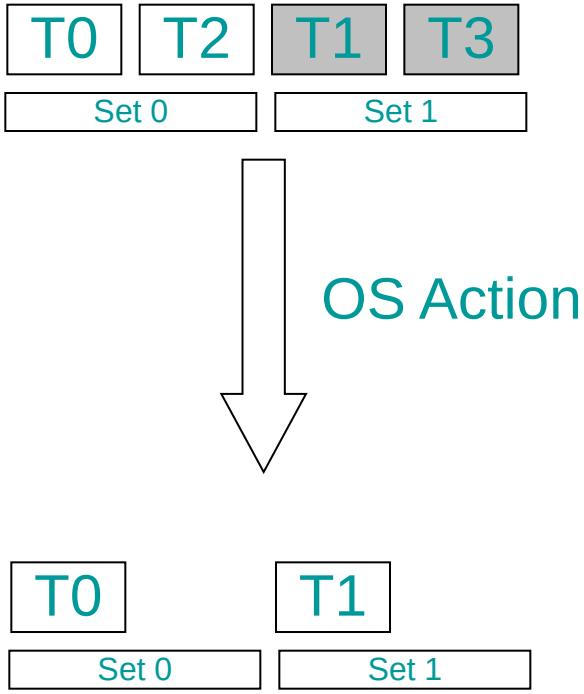
- Augments sampling-based performance analysis and profiling
- Detailed information is collected for sampled instruction
  - Instruction type
  - CPI Stack
  - Branch prediction
  - Cache access
  - Translation

Additional Performance Monitor Enhancements

- Sample Filtering
  - “Needle in haystack” problem
  - Reduces number of samples presented to software by filtering out un-interesting ones
- Hotness table
  - Hardware keeps track of recently sampled addresses and generates an interrupt if the address is “hot”
- Branch History Rolling Buffer
  - Rolling list of recent branches
  - Can be used to detect branch prediction problems
  - Can be used as a call trace leading up to Performance Monitor interrupt
- Event-Based Branches (User Mode Interrupts)
  - Allows user-mode programs to catch Performance Monitor alerts
  - Reduces overhead for user-mode programs to monitor themselves

POWER7 SMT Design

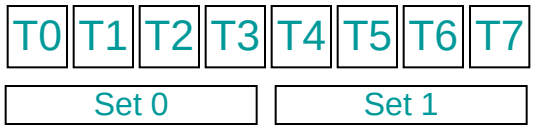




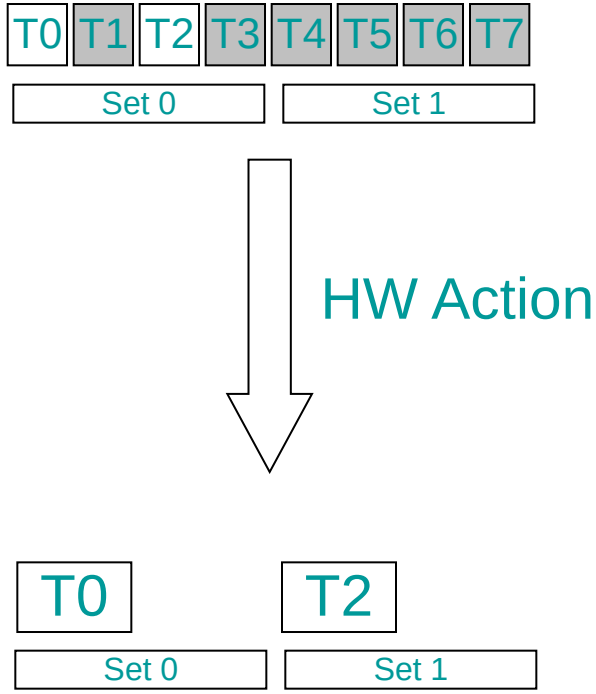
- Divided into two thread sets
  - Static mapping between thread number and thread set
  - Moving to lower SMT level requires
    - Move execution to appropriate thread(s)
    - Nap remaining thread(s)
    - Request SMT level change
  - OS tries to keep threads balanced between thread sets by moving execution to appropriate thread

POWER8 SMT Design

POWER8 automatically tunes itself



- Divided into two thread sets
  - Dynamic mapping between thread number and thread set
  - Moving to lower SMT level requires
    - Nap the idle thread
    - Hardware will shift to the appropriate SMT level
  - Hardware monitors active threads and balances threads between the thread sets



POWER8 Vector/Scalar Unit (VSU)

	POWER7	POWER8
Base SIMD	1X Simple 1X Permute 1X Complex W/DW aligned support	2X Simple (FX and Logical) 2X Permute (byte shuffling manipulation) 2X Complex (integer multiplication) Byte aligned support
Integer SIMD	32 bit integer	64 bit integer 128 bit integer extension/bit permute
Compression /Unstructured data/Parallel Bit Stream Processing	-	On-Chip Accelerator Vector CLZ, Vector Gather bits GR-VR Direct Move
Crypto	-	On-Chip Accelerator AES/SHA User level instructions
RAID CRC/syndrome (Check sum calculation)	-	Vector Polynomial Multiply
Binary Floating Point	8 DP Flops/cyc 8 SP Flops/cyc	8 DP Flops/cyc 16 SP Flops/cyc
Decimal	Non-Pipelined	Pipeline

Hardware Encryption

- On-Chip Hardware Accelerators introduced with POWER7+
  - POWER8 has same accelerators
  - Offload encryption for OS-based large messages (encrypted file systems, etc)
- POWER8 includes user-mode instructions to accelerate common algorithms
- Application Performance Improvements
  - SERT CryptoAES – +133%
  - SPECjbb2013 - +10%

Algorithm	POWER7+	POWER8	
	On-Chip	On-Chip	In-Core
AES-GCM	✓	✓	✓
AES-CTR	✓	✓	✓
AES-CBC	✓	✓	✓
AES-ECB	✓	✓	✓
SHA-256	✓	✓	✓
SHA-512	✓	✓	✓
RNG	✓	✓	
CRC			✓

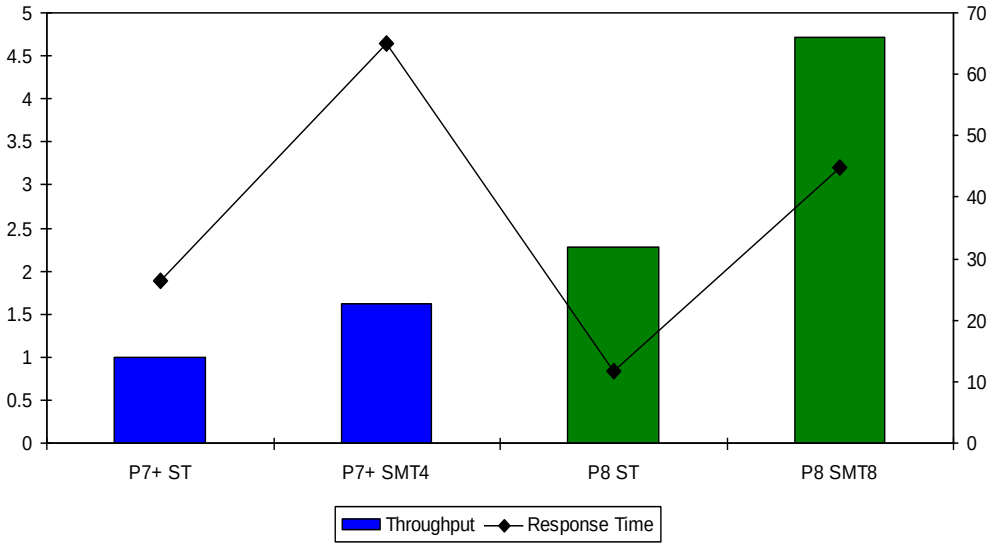
Cycles per Byte

Algorithm	POWER7[+] (SW)	POWER8 (HW)	
		Single Thread	Multi Thread
SHA512	35	10.7	2.6
AES-128-ENC	17	4	0.8
AES-256-ENC	21	5.5	1.1

POWER8 Batch Performance

POWER8 Reduces Batch Window Requirements

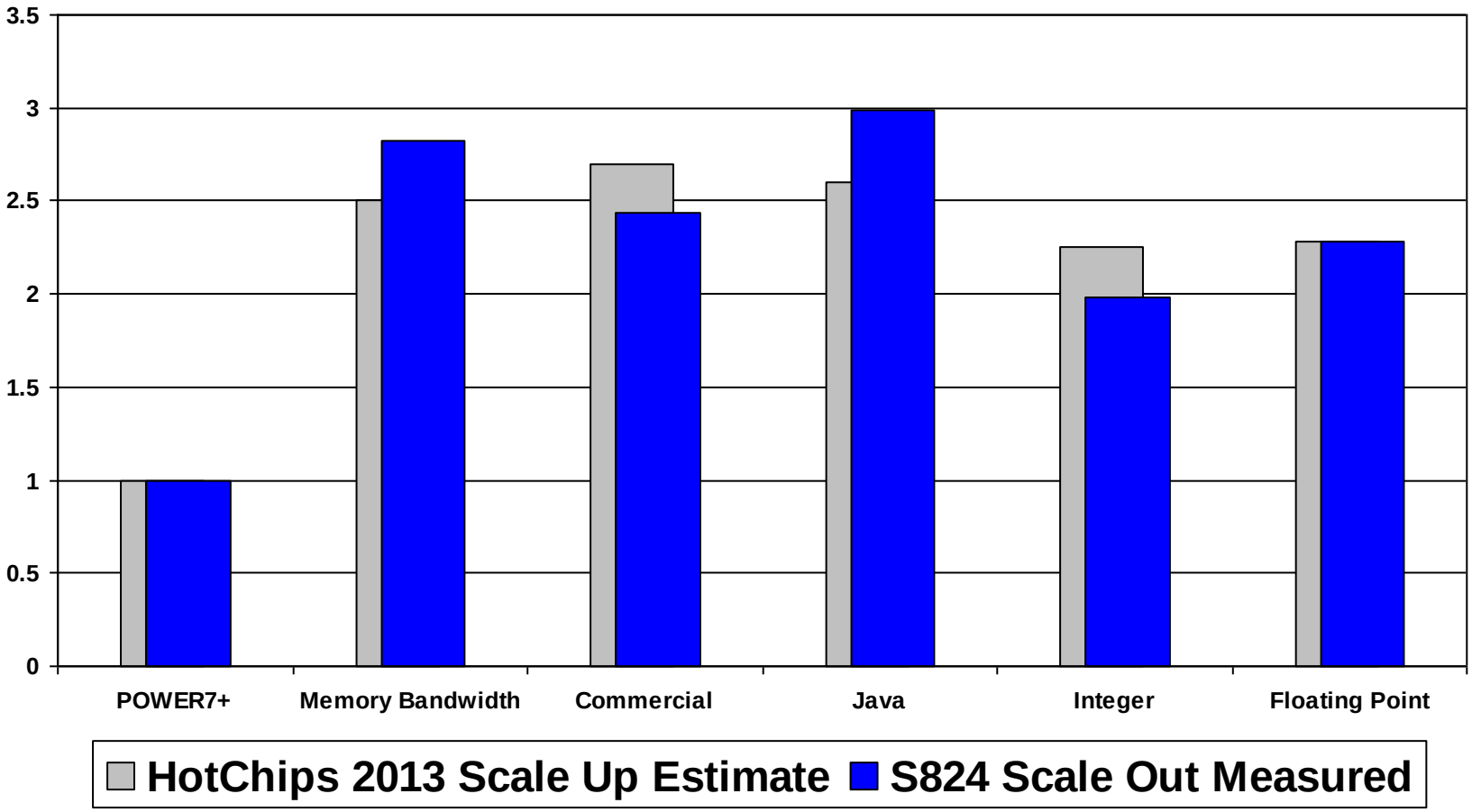
- **56% lower response time and 2.3x more throughput** with POWER8 (Single Thread mode) than POWER7+ (Single Thread Mode)
- **82% lower response time and 1.4x more throughput** with POWER8 (Single Thread mode) than POWER7+ (SMT4)
- **31% lower response time and 2.9x more throughput** with POWER8 (SMT8) than POWER7+ (SMT4)



POWER8 vs. POWER7+ processor performance on an IBM internal workload that emulates batch tasks performing compression where response time is important.

POWER7+ 740 - 16C
POWER8 S824 - 16C

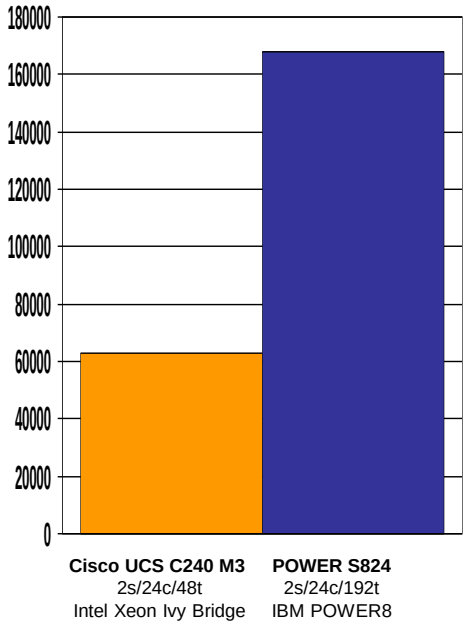
POWER8 Socket Performance



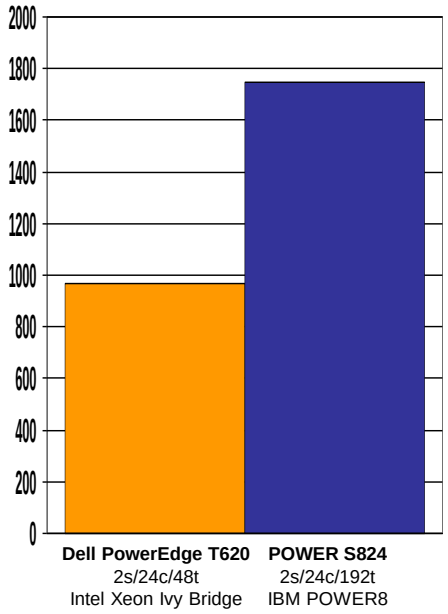
POWER7+ 740 - 16C
POWER8 S824 - 24C

Up to 2.7x performance across key workloads vs. other 24-core Scale-Out Systems

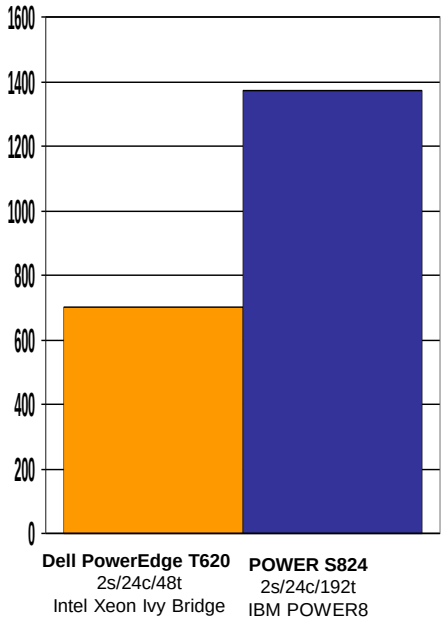
Java – SPECjbb2013 (Max-jOPS)
2.7x Performance



SPECint_rate2006
1.8x Performance



SPECfp_rate2006
2x Performance



1) Results are based on best published results on Xeon E5-2697 v2 from the top 5 Intel system vendors.
2) SPECjbb2013 results are valid as of 7/7/2014. For more information go to http://www.specbench.org/jbb2013/results
3) SPECcpu2006 results are submitted as of 4/22/2014. For more information go to http://www.specbench.org/cpu2006/results/

POWER8

Designed for Big Data - optimized for analytics performance

Processors
flexible, fast execution of analytics algorithms

Memory
large, fast workspace to maximize business insight

Data Bandwidth
bring massive amounts of information to compute resources in real-time

Optimized for a broad range of data and analytics:

Cognos SPSS

IBM Predictive Customer Intelligence

Industry Solutions

Thank You!

Definitions

- eDRAM = embedded DRAM
- SMP = Simultaneous Multi-Processing
- SMT = Simultaneous Multi-Threading
- SR-IOV = Single Root I/O Virtualization
- HMC = Hardware Management Console
- SFF = Small Form Factor

Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.

Special notices (cont.)